动 物 学 研 究 1999, Feb. 20 (1): 21～25 CN 53-1040/Q ISSN 0254-5853
Zoological Research

蛋白质序列和结构的保守性与其功能的关系*

黄京飞[①②] Tom L. Blundell[②]
(①中国科学院昆明动物研究所细胞与分子进化开放研究实验室 昆明 650223)
(②剑桥大学生物化学系 英国剑桥 CB2、1GA)

摘 要 以蛋白质酪氨酸磷酸酶为模型，通过序列和结构的比较，对蛋白质序列、结构的保守性与其功能和进化的关系问题进行了研究。结果显示，虽然在酪氨酸磷酸酶超家族分子的序列中，仅有3个与其催化功能密切相关的残基是高度保守的，但它们功能结构域的核心拓扑结构却明显类似，其中存在着 βαβ 和 βαβα 2个保守的结构单元；此外，它们活性位点的拓扑结构也极其相似。因此，在保持蛋白质功能方面具有重要作用的残基是高度保守的，而具维持蛋白质结构作用的残基则是相对保守的。在进化过程中，蛋白质三维拓扑结构的保守性，似乎主要是体现在保持某些共同的特有二级结构单元和共同的折叠方式上。

关键词 酪氨酸磷酸酶，保守性，功能、进化
中图分类号 Q51

蛋白质序列-结构-功能的关系问题是结构分子生物学的核心问题之一。大量研究表明，虽然在同源蛋白质家族中，相似序列的蛋白质具有相似的结构和功能，但是，在许多蛋白质超家族中，序列差异很大的蛋白质分子却也具有明显类似的结构和功能，这使得单纯的“序列决定结构，结构决定功能”的传统观念正受到严峻地挑战。许多研究显示，蛋白质序列及其结构的保守性，既是其分子功能约束的结果，也是其分子进化的结果。为了探讨在分子进化过程中，蛋白质序列-结构-功能的关系，有必要就蛋白质序列与结构的保守性做进一步地分析研究。

为此目的，我们以蛋白质酪氨酸磷酸酶（protein tyrosine phosphatase, PTP）作为模型，通过对其序列和结构的比较研究，结合有关生物化学和分子生物学的研究结果，讨论了PTP序列和结构的保守性与其功能的关系，旨在为进一步研究蛋白质序列-结构-功能的关系提供一些启示和信息。

1 材料与方法

本研究所涉及的PTP序列及三维结构数据，全部取自Brookhaven的蛋白质数据库（PDB）。序列和结构的比较主要针对功能结构域进行。结构域的划分用DIAL程序（Sowdhamini等，1995）进行。在通过用MNYFIT程序（Sali等，1990）对PTP的功能结构域进行结构重叠后，再运用COMPARER程序（Sutcliffe等，1987）进行基于结构的序列联配。为了能对PTP整个功能结构域的序列进行联配，序列中的一些较大的插入已被删除。结构重叠图通过运用RasMol程序绘制，联配后的序列用JOY程序（Overington等，1990，1992）进行表示。

2 结 果

1phr、2hnq、1ypt和1vhr的结构比较表明，在1phr核心结构中的β-片层结构的顺序不同于2hnq、1ypt和1vhr的。1phr核心结构中，含4个平行的β-片层结构片段，其顺序为：β2-β1-β3-β4；其余3种PTP则有5个与此顺序不同的β-片层结构片段，并且其中有1个β-片层结构片段与其他的4个反平行，而此反平行的β-片层结构片段在1phr却不存在。

1phr活性位点氨基酸顺序的位置与2hnq、1ypt和1vhr的有很大差异，1phr的活性位点片段

* 王宽诚皇家学会奖学金、中国科学院留学经费择优支持基金、云南省应用基础研究基金委主任基金、中国科学院昆明动物研究所细胞与分子进化开放研究实验室基金资助项目
本文1998-03-31收到、1998-10-09修回

($C-X_5-R$) 位于其N-末端，而其余3种磷酸酶的则位于C-末端。然而，二者的活性位点及含Asp环的基本拓扑结构却是相同的。为了使1phr的序列能够与2hnq、1ypt和1vhr的联配，我们将1phr N-端的1～34残基切下，并将其接至C-末端，经过这样的处理，我们对这4种磷酸酶进行了结构重叠（图1）并得到了它们的序列联配图（图2）。

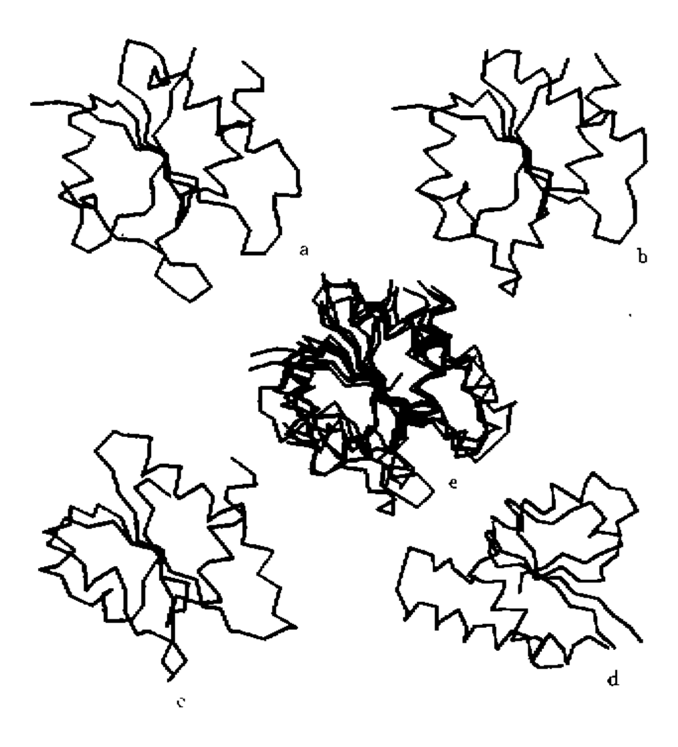


图1 酪氨酸磷酸酶结构的核心区域及其重叠比较
Fig. 1 Structures of core regions and superposition in protein tyrosine phosphatases

a: 2hnq核心区域的结构（2hnq core structure）；b: 1ypt核心区域的结构（1ypt core structure）；c: 1vhr核心区域的结构（1vhr core structure）；d: 1phr核心区域的结构（1phr core structure）；e: 酪氨酸磷酸酶结构核心区域的重叠比较（superposition of core regions in protein tyrosine phosphatases）。

虽然1phr与2hnq、1ypt和1vhr的序列一致性很低（<15%），但它们却保留着1个共同的$C-X_5-R$特征活性位点片段和1个保守的Asp残基（即1phr的Asp129，它相应于2hnq的Asp181）（图2）。已有研究表明，这3个保守的残基都与磷酸酶的功能密切相关（Stone等，1994；Su等，1994；Barford等，1995）。1phr与2hnq、1ypt及1vhr的三维拓扑结构也很类似（图1），在50个等价Cα-原子的结构重叠部分，它们之间的均方根（RMS）误差分别为：1.65×10^{-10} m、1.45×10^{-10} m和1.53×10^{-10} m，并且在核心结构中有βαβ和βαβα 2个保守的结构单元（图2）。1phr与2hnq、1ypt和1vhr等价的绝大多数残基分布于这些保守的β-和α-结构片段中。此外，不仅1phr中与磷酸相结合的环的拓扑结构与2hnq、1ypt和1vhr的非常相似，而且它们活性中心的拓扑结构也能很好地重叠（图3）。

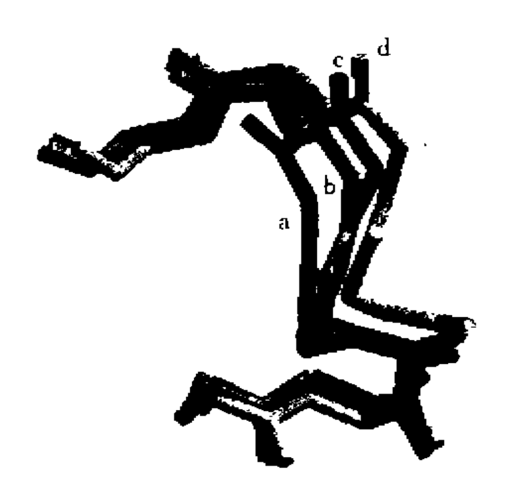


图3 酪氨酸磷酸酶活性位点结构的重叠比较
Fig. 3 The view of the protein tyrosine phosphatase active site superposition

a: 2hnq; b: 1ypt; c: 1vhr; d: 1phr。

3 讨 论

蛋白质中酪氨酸的磷酸化与去磷酸化，在调控细胞的分裂、分化和发育等信号传递过程中，具有重要的作用。磷酸化与去磷酸化作用分别由蛋白质激酶和磷酸酶来催化，PTP则是催化其蛋白质底物中磷酸化的酪氨酸去磷酸化。在催化水解磷酸单酯的过程中，PTP以1个Cys残基作为亲核试剂，并且形成一个磷-Cys的中间产物（Guan等，1991）。通常，PTP被分成两个超家族，即蛋白质磷酸酶Ⅰ和Ⅱ。磷酸酶Ⅰ仅具有一个低分子量的磷酪氨酸磷酸酶（lmwPTP）家族；而磷酸酶Ⅱ则包括两个家族，一个是高分子量磷酪氨酸磷酸酶，另一个为双重专一性磷酸酶。截止1997年底，在PDB中已收录三维晶体结构数据的PTP共有4种，即lmwPTP（1phr）、PTP1B（2hnq）、*Yersinia* PTP（1ypt）和双重专一性磷酸酶（1vhr）。这4种酶中，2hnq和1ypt属于高分子量磷酪氨酸磷酸酶家族，1vhr属双重专一性磷酸酶家族，这3种酶同

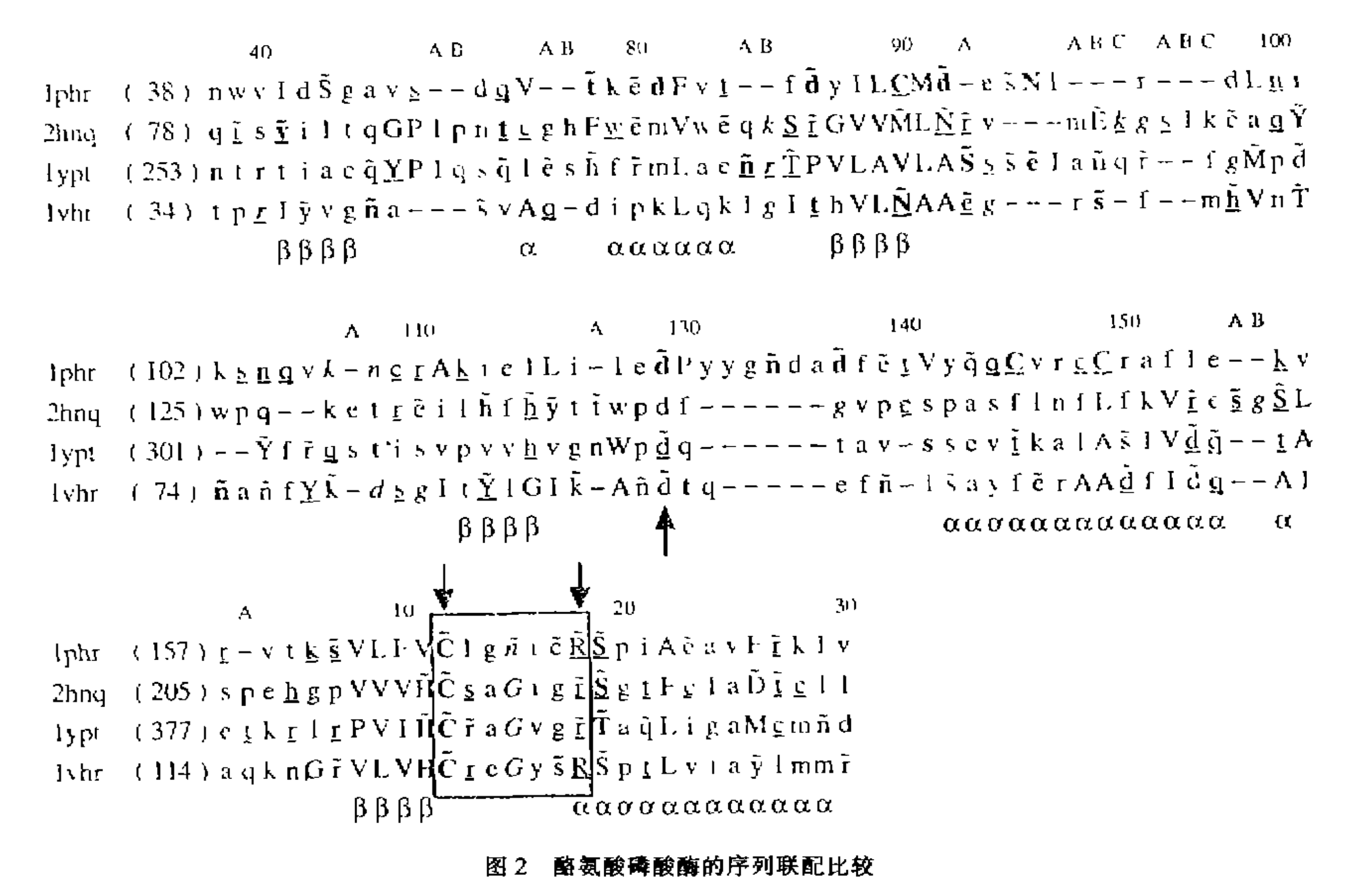


图2　酪氨酸磷酸酶的序列联配比较

Fig. 2　The sequence alignment of protein tyrosine phosphatases

方框示：磷酸酶的特征序列片段；箭头示：保守残基；α：α－螺旋片段；β：β－片层片段；序列中各种符号所表示残基的不同构象及性质分别为，X：溶剂不可及的残基；x：溶剂可及的残基；*x*：正φ构象角残基；x̃：与其他侧链成氢键的残基；**x**：与主链的酰氨成氢键的残基；x̲：与主链的羧基成氢键的残基［box：signature motif；arrows：conserved residues；α：α-helix elements；β：β-sheet elements；residues in the alignment are represented as follows，X：solvent inaccessible；x：solvent accessible；*x*：positiveφ；x̃：hydrogen bond to other sidechain；**x**：hydrogen bond to mainchain amide；x̲：hydrogen bond to mainchain carbonyl］。

属磷酸酶Ⅱ超家族，而1phr则属磷酸酶Ⅰ超家族。

1phr和2hnq、1ypt及1vhr不仅具有类似的三维结构、共同的特征性活性位点片段和一个具催化活性的保守的Asp残基，而且它们还具有相同的反应催化机制，包括相同的磷－Cys中间产物和催化反应底物等，因而，磷酸酶Ⅰ与Ⅱ应属于同一蛋白质超家族，至于它们序列间的较大差异，则可能是长期进化的结果。

蛋白质序列、结构与功能间的关系，至今仍不完全清楚，一方面，序列高度一致的蛋白质，其结构和功能高度相似；另一方面，一些序列一致性较低的蛋白质却也具有类似的结构和功能，如2hnq、1ypt、1vhr和1phr。因此，蛋白质序列、结构与功能的关系并不是一种线性的关系。目前，PDB中已有三维结构数据的蛋白质数目已超过7 000，而至今尚未发现有两种不同蛋白质分子的序列是完全一致的。然而，通过对蛋白质结构的分类比较研究发现，所有已知三维结构的蛋白质，大致可分为全α、全β、α+β和α/β 4个大类、327种折叠方式、463个超家族和652个家族（Brenner等，1997）。毫无疑问，蛋白质的结构类型数远远低于它的种类数。一般来讲，同一家族中的蛋白质功能是相同的，甚至有些同一超家族中的蛋白质功能也是相同的。而目前已知三维结构的所有蛋白质仅652个家族，显然大大少于蛋白质的种类数。因此，在分子进化过程中，不仅蛋白质的三维结构较其序列更为保守（Bajaj等，1984），实际上，蛋白质的功能也是高度保守的。然而，从某种意义上讲，结构与功能间的关系，似乎没有功能与序列中若干重要残基的关系那么严格。许多研究显示，仅仅1个残基的替

换，就可能导致蛋白质功能的完全丧失或显著降低；而序列差异很大的不同蛋白质分子，则既可具有非常类似的结构和相同的功能，也可具有功能相差甚远的某种相似结构。一种类似的结构单元可以由不同的序列片段构成，例如，磷酸酶中由不同氨基酸顺序组成的βαβ结构单元，也存在于脱氢酶、黄氧还蛋白和激酶等蛋白质中。因此，蛋白质的三维结构也许主要是受物理、化学和几何因素的约束，而不是主要受其序列的约束。

在进化过程中，维持蛋白质结构的残基不如保持蛋白质功能的残基保守（黄京飞等，1998）。目前，几乎所有关于序列与结构比较的研究都支持这样的观点：对于功能重要的残基是保守的，并且位于拓扑结构的等价位置（Overington 等，1990，1992）。然而，蛋白质三维拓扑结构的保守性，则似乎主要体现在保持某些共同的特有二级结构单元和折叠方式上。例如来自226种球蛋白序列和结构的比较表明，虽然它们的三维拓扑结构非常相似，但其中有的序列的一致性却只有16%，而且在所有的序列中仅有2个残基是保守的，但在进化中几乎所有的α-螺旋结构片段都得以保存下来（Bashford 等，1987）；在PTP中，也只有3个残基是保守的，但其核心结构中却保留了βαβ和βαβα 2个保守的结构单元，尽管组成这2个结构单元的残基不尽相同。因此，我们认为，蛋白质的拓扑结构对于保持蛋白质的功能是重要的，不同的拓扑结构将导致不同的功能。虽然拓扑结构对于保持功能是必要的，但仅只是拓扑结构还不足以保证蛋白质功能的实现。在满足某种蛋白质特定功能所要求的拓扑结构的前提下，一些关键残基的存在，将是蛋白质功能得以实现的重要因素。因此，对于保持蛋白质功能非常重要的残基是高度保守的。但是，维持蛋白质结构的残基则是相对保守的，因为它们可以在蛋白质结构的任何位置上，从几个不同的"候选"残基中选择其中之一。

致 谢 本工作承蒙英国剑桥大学生物化学系 R. Sowdhamini 博士、D. Burke 博士及 N. Srinivasan 博士予以协助，特此致谢。

参 考 文 献

黄京飞，刘次全，1998. 氨基酸残基可及性与蛋白质家族成员结构的保守性. 动物学研究，19(2)：137～142. [Huang J-F, Liu C-Q, 1998. Amino acid residue accessibilities and structural conservativeness of molecules in protein family. *Zool. Res.*, 19(2): 137－142.]

Bajaj M, Blundell T, 1984 Evolution and the tertiary structure of proteins. *Ann. Rev. Biophys Bioeng.*, 13: 453－492

Barford D, Jia Z, Tonks N K, 1995. Protein tyrosine phosphatases take off. *Nature Struct. Biol.*, 12: 1043－1053.

Bashford D, Chothia C, Lesk A M, 1987. Determinants of a protein fold. Unique features of the globin amino acid sequences. *J. Mol. Biol.*, 196: 199－216.

Brenner S E, Chothia C, Hubbard T J P, 1997. Population statistics of protein structure: lessons from structural classifications. *Current Opinion Struct. Biol.*, 7: 369－376.

Guan K L, Dixon J E, 1991. Evidence of protein-tyrosine-phosphatase catalysis proceding *via* a cysteine-phosphate intermediate. *J. Biol. Chem.*, 266: 17206－17030.

Overington J P, Johnson M S, Sali A *et al*, 1990. Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction, *Proc. R. Roc. Lond.*, B241: 132－145.

Overington J P, Donnelly D, Sali A *et al*, 1992. Environment specific amino acid substitution tables: tertiary templates and prediction of protein folds. *Protein Sci.*, 1: 216－226.

Sali A, Blundell T L, 1990. The definition of topological equivalence in homologous and analogous structures: a procedure involving a comparison of local properties and relationship. *J. Mol. Biol.*, 212: 403－442.

Sowdhamini R, Blundell T L, 1995. An automatic method involving cluster analysis of secondary structures for the identification of domains in proteins. *Protein Sci.*, 4: 506－520.

Stone R L, Dixon J E, 1994. Protein-tyrosine phosphatases. *J. Biol. Chem.*, 269: 31323－31326.

Su X-D, Taddei, N, Stefani M *et al*, 1994. The crystal structure of a low-molecular-weight phosphotyrosine protein phosphatase. *Nature*, 370: 575－578.

Sutcliffe M J, Hayes F R F, Camet D *et al*, 1987. Knowledge-based modeling of homologous proteins. I. Three-dimensional frameworks derived from simultaneous superposition of multiple structures. *Protein Eng.*, 1: 377－384.

THE RELATIONS OF PROTEIN SEQUENCE AND STRUCTURAL CONSERVATION WITH FUNCTION

HUANG Jing-fei①② Tom L. Blundell②

(①*Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, the Chinese Academy of Sciences, Kunming* 650223)

(②*Department of Biochemistry, University of Cambridge*, 80 *Tennis Court Road, Cambridge* CB2, 1GA, U.K.)

Abstract The conservation in sequence and structure of protein tyrosine phosphatases has been analyzed by aligning their sequences and superposed their three-dimensional topological structures. The results indicated that only three residues related closely to the function are conserved in sequence, but the core regions in their functional domains are strikingly similar in structure, and there are two conserved structural motifs, *i.e.*, βαβ and βαβα motifs. On the other hand, the topology of the active site in protein tyrosine phosphatases is also very similar. Thus, it is suggested that the very important residues for maintaining protein function are highly conserved, however, the residues for keeping protein structure are conservatively varied. In molecular evolution, three-dimensional structural conservation seems to be mainly expressed in holding some common secondary structural elements and common fold.

Key words Protein tyrosine phosphatase, Conservation, Function, Evolution